Appendix A

Distributions Used in Uncertainty Analysis

The Beta distribution

The Beta distribution is one of three distributions associate with a binomial stochastic process. A binomial process is a random counting system where there are a discrete number of opportunities (trials) of some particular event happening (successes) and where each trial has the same probability of being a success. This means that each trial must be independent of every other trial.

There are many systems that closely approximate a binomial process. Random processes like the tossing of a coin are binomial, since one face of the coin can be defined as being a success and the probability of each coin being a success remains constant for all tosses. No matter how many "heads" there have been in a row, the probability of a "tails" for the next toss remains the same (e.g. 50% for a fair coin).

Random sampling from a population may also closely approximate a binomial process, where we are concerned with determining what proportion of that population has some characteristic of interest. If the population is much larger than the sample size (a rule of thumb is that the population should be at least 10 times the size of the sample) then the probability of an individual randomly sampled from the population having the characteristic of interest remains fairly constant and equivalent to the proportion of the population with that characteristic. So, for example, if we are interested in the proportion of US citizens that eat meat, we can do a random survey of US citizens. Providing that our sample is much smaller than the population size, the probability that each consecutive randomly selected person eats meat remains reasonably constant, though there are actually a finite number of people who eat meat and there is a finite population too, which means that the probability that the next randomly samples person eats meat does, in fact, depend on the previous samples.

The example above for meat eaters assumes that we are sampling without replacement: in other words, we would not survey any person more than once. It also implicitly assumes that there are a fixed number of people who eat meat in the population and a fixed population size too. This would be a static system with a constant proportion of meat eaters.

The Beta distribution can be used to model the confidence one has about the probability of success of a binomial trial p where one has observed n independent trials of which s were successes (27, 47, 73, 92, 93), so that it is said that p is distributed as a Beta(s+1,n-s+1).

This distribution is the result of applying Bayes' Theorem with a Uniform(0,1) prior distribution and a binomial likelihood function. In layman's terms, Bayes' Theorem works as follows:

1. A prior statement of the knowledge of the variable to be modelled is given. In this case, we are saying that the probability p lies somewhere between zero and one, but we would not like to say that any value within that range was anymore likely than any other value (hence the Uniform(0,1) prior distribution.
2. For each allowed value within the prior distribution's range, we calculate the probability of observing the s successes we observed from the n trials. This probability is simply the binomial probability:

$$P(s;n,p) =_n C_s p^s (1-p)^{n-s}$$

3. These binomial probabilities then become the weightings given to each value of p in the prior distribution. By normalising these weightings we arrive at a posterior (i.e. final answer) distribution.

Thus, the posterior distribution *f(p)* for *p* is given by the product of the prior distribution density and the likelihood function:

$$f(p;s,n) = \frac{1.{}_nC_s\,p^s(1-p)^{n-s}}{1.\int_0^1 {}_nC_s\,t^s(1-t)^{n-s}\,dt} \qquad 0<p<1$$

where the 1 in the equation is the probability density of the Uniform(0,1) prior distribution, the ${}_nC_s\,p^s(1-p)^{n-s}$ is the weighting given to the *p* value (the likelihood function) and the integral in the denominator normalises the distribution so that the area under the curve equals unity. By omitting the 1, cancelling out the ${}_nC_s$ and using a_1=*s*+1, a_2=*n*-*s*+1, we arrive at the equation for the Beta(a_1, a_2) distribution's probability density $f(x;a_1,a_2)$:

$$f(x;a_1,a_2) = \frac{p^{a_1-1}(1-p)^{a_2-1}}{\int_0^1 t^{a_1-1}(1-t)^{a_2-1}\,dt}$$

Note that the prior distribution for *p* is a Uniform(0,1) distribution, shown in Figure A1. The Uniform(0,1) prior distribution is an uninformed prior, meaning that no subjective opinion or any other information has been involved in determining the prior. This is logically the most conservative approach one could take where conservatism here means expressing the maximum degree of uncertainty possible. The selection of an appropriate prior is sometimes slightly contentious. For example, Beta(0,0) is sometimes suggested as an uninformed prior, though it does not in theory exist. One criticism for using a Beta(1,1) prior is that the mean of the estimated probability is biased towards 50%, away from the observed proportion. In fact, for all applications of the Beta distribution in this analysis, the information contained in the prior distribution is generally overwhelmed by the information contained in the sample data and the results are essentially equivalent to a more traditional frequentist statistics approach which would give the confidence distribution for the probability *p* as: $\hat{p}$ = Binomial(*n*, *s*/*n*) / *n*. The frequentist also uses the central limit theorem in situations where a large number *n* of samples were taken, say *n* > 30. By the central limit theorem $\hat{p}$ has a Normal distribution Normal(*p*, $\sqrt{[p(1-p)/n]}$), where *p* is the true value of the probability of success. One then assumes (i.e. estimates) the confidence distribution for *p* to be :

$$\hat{p} = Normal\left(\frac{s}{n}, \sqrt{\frac{s(n-s)}{n^3}}\right)$$

Thus, from a practical viewpoint, there is little difference in using a Bayesian or frequentist approach to uncertainty estimating in this model, except that the Bayesian approach allows one to combine information from dissimilar data.

Perhaps the easiest way to understand the Beta distribution used in this manner is to look at a few plots of its shape for varying values of a_1 and a_2. Figure A1, the Uniform(0,1) distribution, is also the Beta(1,1), i.e. the Beta distribution where we have observed *s* = 0 successes and (*n*-*s*) = 0 failures: this is the distribution when we have not yet done any trials and hence remains the prior distribution.

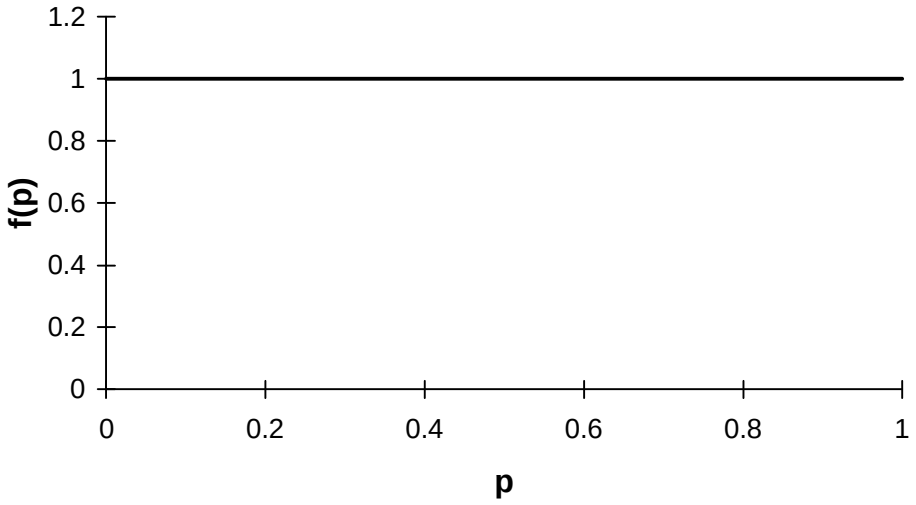

Figure A1: The Uniform(0,1) distribution, which is also the Beta(1,1) distribution

Figure A2 show the Beta(1,11)and the Beta(11,1) distributions: the former where we have observed zero successes in ten trials and the latter where we have observed ten successes in ten trials. Note that they are simply the reflection of each other since they essentially represent the same thing: one needs only to reverse the definition of a success to its opposite. Also note that, since all trials have been a success or a failure, the distributions peak at zero and one respectively. If this pattern continues with more trials, the distributions will become progressively more concentrated at zero and one.

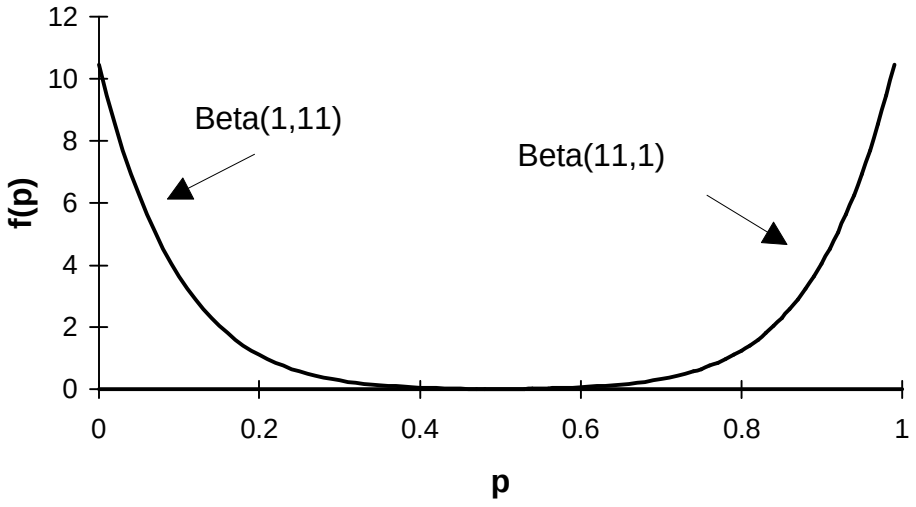

Figure A2: Examples of the Beta distribution where all trials are successes (peak at p=1) or all are failures (peak at p=0)

Figure A3 shows the Beta(3,3), Beta(11,11) and Beta(21,21) distributions representing four, 20 and 40 trials where 50% have been successes. Note that as the number of trials increases, the distribution becomes progressively narrower: in other words, one is becoming progressively more confident about what the true value of p must be.

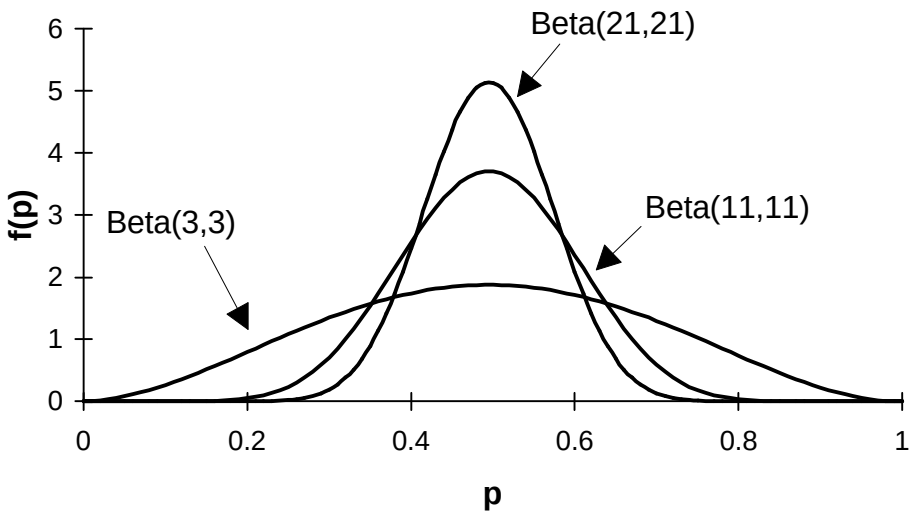

Figure A3: Examples of the Beta distribution where there are equal successes and failures (i.e. $a_1=a_2$)

The figures below show that except for the case with very small *n*, the Binomial distributions or the Normal approximations to them that a frequentist would use to model the confidence distributions modelled as Beta posterior distributions above are practically indistinguishable from the Beta distributions. The cumulative probability distributions for the Binomial(4,0.5)/4, the Normal(0.5, 0.25), and the Beta(3,3) which are the ones for the case n=4, are shown in Figure A4a. The cumulative distribution for the Binomial is a step function. Figure A4b displays the cumulative probability distributions for the case n=20 and Figure A4c, for the case n=40.

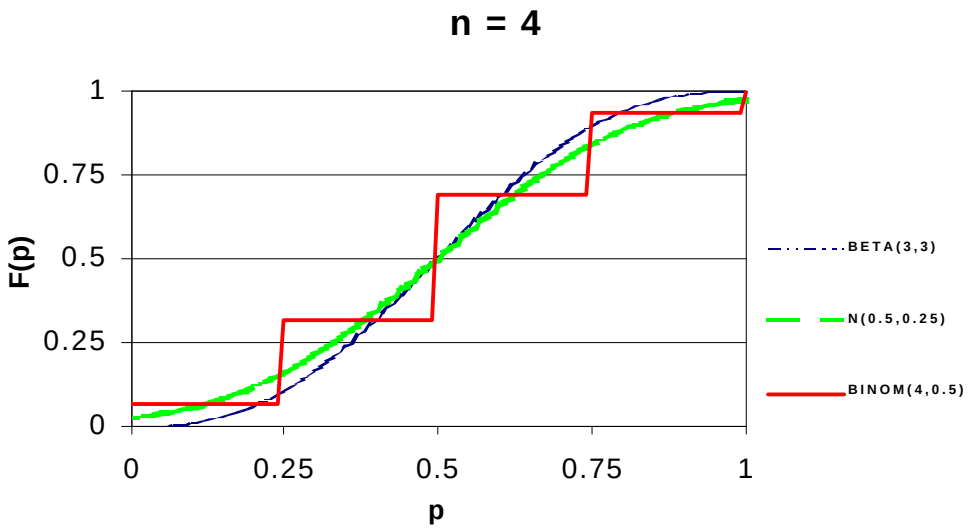

Figure A4a. Cumulative probability distributions for *p* based on the Beta, Binomial and Normal for the case n=4 and the number of observed successes is 2.

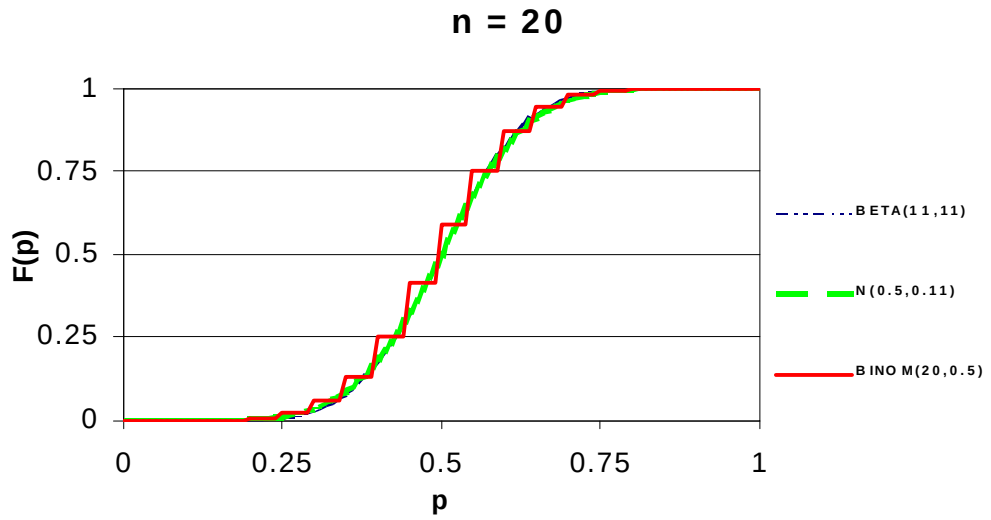

Figure A4b. Cumulative probability distributions for *p* based on the Beta, Binomial and Normal for the case n=20 and the number of observed successes is 10.

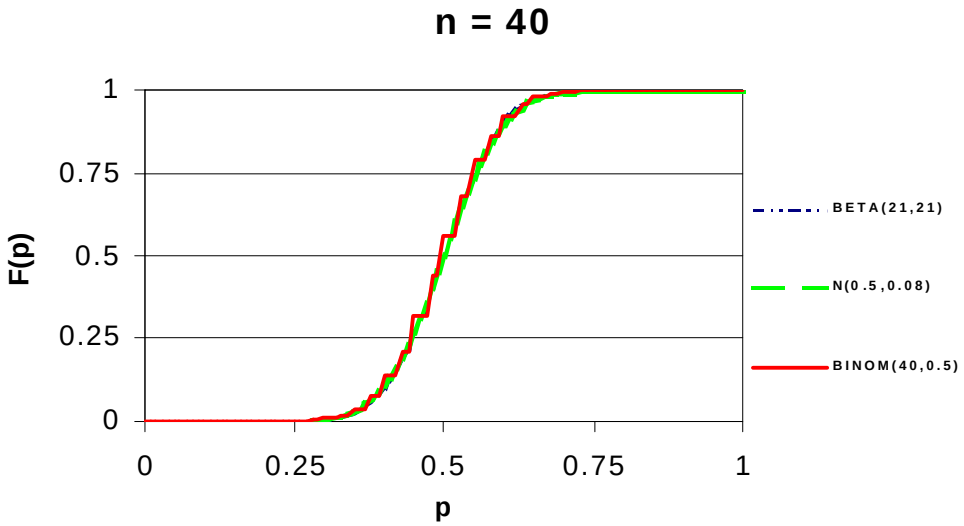

Figure A4c. Cumulative probability distributions for *p* based on the Beta, Binomial and Normal for the case n=40 and the number of observed successes is 20.

Use of the Gamma distribution to describe the uncertainty about a Poisson mean

Like the binomial probability *p*, the mean events per period l is a fundamental property of the stochastic system in question. It can never be observed and it can never be exactly known. However, we can become progressively more certain about its value as more data are collected. Bayesian inference again provides us with a means of quantifying the state of our knowledge as we accumulate data.

Let us assume an uninformed prior $p(l) = 1/l$. The Poisson likelihood function for observing *X* events in period *t* is given by:

$$l(X \mid l, t) = \frac{e^{-lt}(lt)^X}{X!} \propto e^{-lt}(l)^X$$

The posterior distribution density is the product of the prior density and the likelihood function. We can ignore terms that don't involve l, recognising that the distribution will be normalised eventually, and we then get the posterior distribution:

$$p(l \mid X) \propto e^{-lt} l^{X-1} \qquad l > 0$$

which has the functional form of, and therefore is, a Gamma(X,1/t) distribution where l is the variable.

The shape of the posterior Gamma distribution becomes progressively less sensitive to the prior distribution as data is collected. In essence, the sensitivity of the Gamma distribution to the prior amounts to whether (X-1) is approximately the same as X. So, if X was 100, the difference would be roughly 1% influenced by the prior and 99% influenced by the data. In this model, the information contained in the quantity of data available *always* overpowers the prior.

Gamma distributions used in the risk assessment to model the rate of invasive infection, λ_i which has a Gamma(43,1) distribution and the rate of enteric infection, λ_e which has a Gamma(3985,1) distribution. Those distributions are shown in the following graphs.

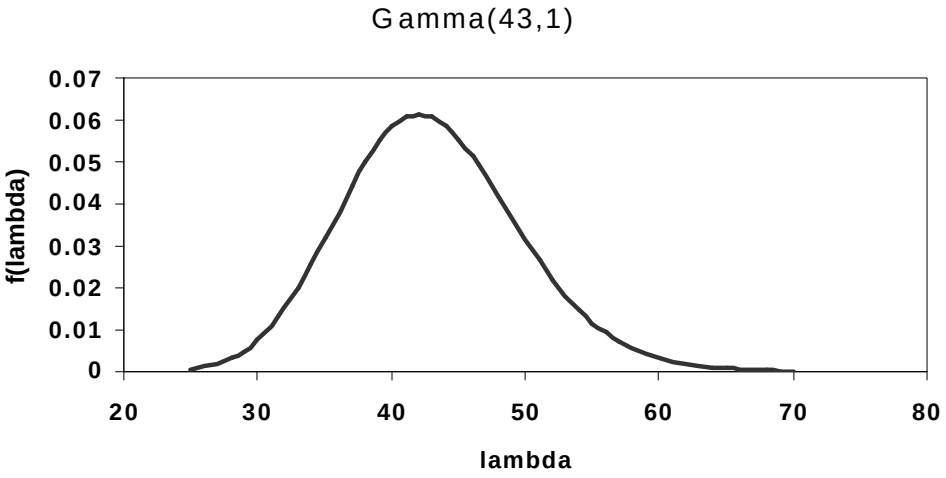

Figure A5a. The Gamma(43,1) distribution used to model the rate of invasive disease, λ_i.

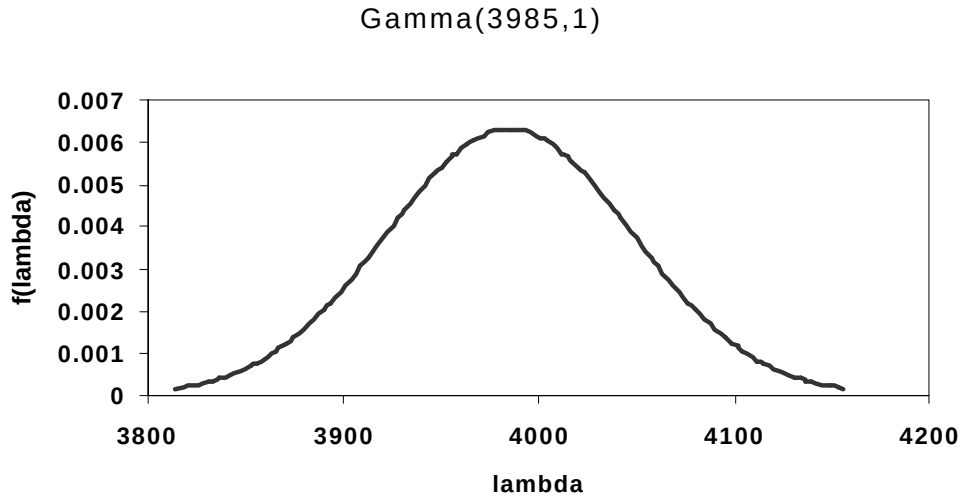

Figure A5b. The Gamma(3985,1) distribution used to model the rate of enteric disease, λ_e.

As was the case for the Beta distribution, there is a distributional counterpart to the gamma that the frequentist would apply to this estimation problem. When X, the number of events in one unit of time (t=1), is distributed Poisson(λ),the maximum likelihood estimator for λ, $\hat{l}$ is the observed value of X. Then the Poisson(X) distribution is used as the uncertainty distribution for λ. The Poisson distribution is the limiting distribution for a Binomial random variable when *n* is large and *p* is very small. Because of this relationship, the central limit theorem applies for the Poisson under the same limiting conditions. By the central limit theorem $\hat{l}$ is normal with mean λ and variance λ, where λ is the true value of the parameter. This means that the confidence distribution for λ is estimated as:

$$Normal(X, \sqrt{X})$$

The three possible choices for the confidence distribution for λ when 43 cases were observed during the year are shown to demonstrate how similar they are.

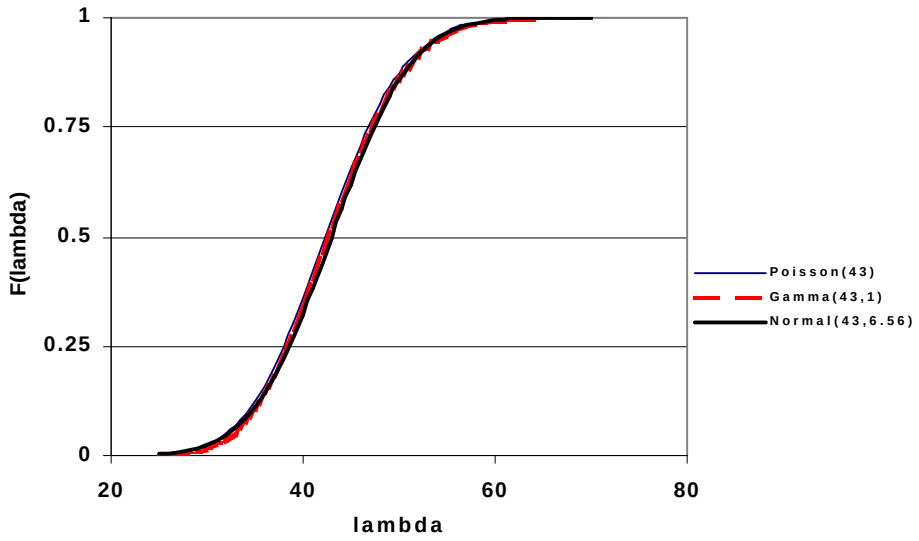

Figure A6. The cumulative distribution functions for Gamma(43,1), Poisson(43), and Normal(43,6.56).

Appendix B

	A	B	C	D	E
1	**Analysis of the human health effect of fluoroquinolone resistant *Campylobacter***				
2	**in domestically reared and consumed broilers**				
3	Key: data, assumption, calculation, *link*, model output, section result				
4	**Section 1**	**Nominal observable confirmed cases of campylobacteriosis in US**			
5	n_{US}	US population	270,298,524		
6	n_{FN}	Catchment site population	20,723,982		
7	o_i	Observed FoodNet invasive cases of campylobacteriosis	43		
8	o_e	Observed FoodNet enteric cases of campylobacteriosis	3,985		
9	I_i	Expected observed invasive disease in catchment	43		
10	I_e	Expected observed enteric disease in catchment	3,985		
11	N_i	Nominal observable mean population invasive infections	561		
12	N_e	Nominal observable mean population enteric infections	51,976		
13	p_b	Propn enteric infections w bloody diarrhea	46.0%		
14			**Enteric**		**Invasive**
15			**Non-bloody**	**Bloody**	
16	$N1_{en}$, $N1_{eb}$, $N1_i$	Nominal mean Culture Confirmed Cases reportable to health department	28,077	23,898	*561*
17	**Section 2**	**Total nominal expected number of *Campylobacter* infections in a year in US**			
18	p_{nm}, p_{bm}	P(seek care)	12%	26.7%	100%
19	p_{nc}, p_{bc}	P(stool requested and submitted)	19%	55.4%	100%
20	p_t	P(lab tests for organism)	94.5%	*94.5%*	100%
21	p_+	P(culture confirmed given tested) = test Se, assumes Sp=1	75.0%	*75.0%*	100%
22	$N2_{en}$, $N2_{eb}$, $N2_i$	Illness in population	1,702,043	228,040	561
23					
24	$N2_T=N2_{en}+N2_{eb}+N2_i$	Total cases (bloody+non-bloody+invasive)	1,930,644		
25	**Section 3**	**Number of fluoroquinolone resistant infections, from domestically reared and consumed chickens**			
26	$p_{ca\text{-}min}$	Lower bound estimate	48.0%		
27	$p_{ca\text{-}max}$	Upper bound estimate	70.0%		
28	p_{ca}	Therefore chicken associated	59.0%		
29		*Chicken associated cases*	1,004,205	134,543	331
30	p_{nm}, p_{bm}	Proportion seeking care	*12.2%*	*26.7%*	*100%*
31		*Number seeking care*	122,078	35,878	331
32	p_{an}, p_{ab}, p_{ai}	Proportion treated with antibiotic	47.9%	63.7%	100%
33		*Number treated*	58,450	22,854	331
34			z	39.5%	
35	p_{FQ}	Proportion receiving FQ treatment	55.08%		*55.08%*
36		*Number of chicken related cases treated with FQ*	32,195	12,588	182
37	p_{rh}	Proportion of *Campylobacter* infections from chicken that are FQ resistant	10.4%		
38	$N3_{en}$, $N3_{eb}$, $N3_i$	Number of fluoroquinolone resistant infections from chicken seeking care, getting fluoroquinolone	3,352	1,311	19
39	***$N3_T=N3_{en}+N3_{eb}+N3_i$***	**Total number of fluoroquinolone resistant infections from chicken seeking care, getting fluoroquinolone**	**4,682**		
40	**Section 4**	**Number of fluoroquinolone-resistant *Campylobacter* contaminated chicken carcasses consumed annually**			
41	p_c	Total prevalence of *Campylobacter*	88.1%		
42	p_{rc}	Prevalence of fluoroquinolone resistant *Campylobacter* among *Campylobacter* isolates from slaughter plant	11.8%		
43	p_p	Estimated prevalence of fluoroquinolone-resistant *Campylobacter* in broiler carcasses	10.4%		
44	c	Consumption of boneless domestically reared chickens in US per head (lbs)	51.40		
45	V_c	Total consumption of boneless domestically reared chicken in US (lbs)	1.39E+10		
46	V_i	**Total consumption of boneless domestically reared chicken contaminated at slaughter plant with fluoroquinolone resistant *Campylobacter* in US (lbs)**	**1.45E+09**		
47		**Denominators**	**Value**	**Probability**	**Equated to 1 in:**
48	*P1*	**US citizen**	**270,298,524**	**0.0017%**	**57,737**
49	*P2*	**Person with campylobacteriosis**	**1,930,644**	**0.2425%**	**412**
50	*P3*	**Person with campylobacteriosis seeking care**	**268,284**	**1.7450%**	**57**
51	*P4*	**Person with campylobacteriosis seeking care and prescribed antibiotic**	**138,364**	**3.3835%**	**30**

Block 2. List of the formulae used to produce the model.

	A	B	C	D	E	F
53	**Formulae table**					
54	Cell	Formula	Cell	Formula		
55	D12:D13	=RiskGamma(D10,1)	D41:E41	=(D24*'Estimates 1'!$I13)+(1-D24)*$E43		
56	D14:D15	=D12*D$8/D$9	E43	=RiskBeta(17,26)		
57	D16	='Estimates 1'!P16	D44	='Estimates 1'!J13		
58	D20	=D15*(1-D16)	F44	=D44		
59	E20	=D15*D16	D45:E45	=D42*$D44		
60	F20	=D14	D47	='Estimates 1'!G25		
61	D23	='Estimates 2'!H13	D48:F48	=$D47*D45		
62	E23	=D15*D16	D49	=SUM(D48:F48)		
63	D24	='Estimates 2'!H16	D52	=RiskBeta(1144+1,1297-1144+1)		
64	E24	='Estimates 2'!E16	D53	=RiskBeta(18+1,159-18+1)		
65	D25	=RiskBeta(367846+1,389255-367846+1)	D54	=D53*D52		
66	D26	=RiskBeta(12,4)	D56	=D55*D8		
67	E25:E26	=D25	D57	=D56*D54		
68	D27:F27	=D20/PRODUCT(D23:D26)	D60	=D8		
69	D29	=D27+E27+F27	D61	=D29		
70	D35	=RiskUniform(D33,D34)	D62	=SUMPRODUCT(D27:F27,D38:F38)		
71	D36:F36	=D27*D35	D63	=SUMPRODUCT(D27:F27,D38:F38,D41:F41)		
72	D38:F38	=D23	E60:E63	=D$49/D60		
73	D39:F39, D42:F42, F45	=D38*D36	F60:F63	=1/E60		